Action-Oriented Energy Benchmarking for Nonresidential Buildings

The paper is concerned with building energy-efficiency benchmarking. Traditional benchmarking addresses the status quo, e.g., by comparing the building to its peers at one point in time or longitudinally. Action-oriented benchmarking extends this process by also inferring potential energy-efficiency opportunities.

By Evan Mills

ABSTRACT | The complex process of improving the energy efficiency of a building begins with understanding baseline conditions and assessing the potential for specific improvements. Traditional benchmarking typically addresses the *status quo*, e.g., by comparing the building to its peers at one point in time or longitudinally. Action-oriented benchmarking extends this process by also inferring potential energy-efficiency opportunities. Doing so, however, requires more in-depth benchmarking than offered by traditional "whole-building" assessment methods. The process begins by carefully identifying a peer group for comparison that has true relevance to the subject building, and then disaggregating energy use by fuels and end uses to better pinpoint inefficiencies. Toward this end, the benchmarking process can be extended from energy to emissions and costs. Building characteristics and energy utilization parameters, as distinct from resource utilization data, can also be benchmarked in order to ascertain potential relevance and applicability of energy-efficient technologies or practices. To ensure savings attainment and persistence, benchmarking must continue throughout a building's lifecycle. A publicly funded web-based benchmarking system called EnergyIQ is introduced, which implements the aforementioned principals.

KEYWORDS | Energy benchmarking; energy efficiency; nonresidential buildings

I. ROLE OF BENCHMARKING IN THE BUILDING ENERGY LIFECYCLE

Benchmarking is a widely used approach to putting information in context, and has long been employed to characterize everything from stock prices to the weather to human intelligence. The use of energy benchmarking has come into practice relatively recently. Methodologies exist for the industrial, transportation, and buildings sectors [7], [11]. The range of technical methodologies has been reviewed elsewhere [3], [4], [12]). This paper focuses on conceptual approaches, best practices, user needs, and the evolving market context for benchmarking with a focus on nonresidential buildings. Techniques for making benchmarking results more "actionable" are also presented.

Energy benchmarking for buildings has come to be recognized as integral to managing energy use across a building's entire lifecycle. Benchmarking is a powerful way to educate and inspire building occupants and other decision makers seeking to design new buildings, improve the performance of existing ones, and ensure persistence of energy savings over time. Benchmarks are also valuable to those targeting and evaluating energy-efficiency programs, policies, or recognition campaigns. By enhancing the transparency of the energy management process, benchmarking plays an important role in reducing real and perceived risks in market transactions that depend on the comparative valuation of energy use and savings.

Prior to initiating the building design process, benchmarking the performance of average and exemplary buildings can be used to inform design intent with respect to aspirational energy savings. Benchmarking can help identify best practice technologies and

operational procedures by delving into the design choices and performance outcomes in exemplary buildings. Once design has commenced, model-based benchmarking can be used to compare simulated performance of the subject building to peer groups or specific targets. These benchmarks can continue to be used once a building is operational (longitudinal benchmarking) to verify attainment of design objectives, diagnose performance problems, and track progress towards performance improvements over time and ensure persistence of savings achieved by physical or operational changes intended to save energy. The identification of key metrics also informs a facility's metering and submetering plan, fault- and energy-anomaly detection protocols, target setting. For all use cases, outliers can be studied to identify best practices as well as critical causes of energy inefficiencies.

The benchmarking process can take many forms, and be performed at various scales. At one extreme, national aggregate average building energy intensities from one country could be compared to those of others, while at the other extreme a single end use (e.g., lighting) in an individual building could be benchmarked against those of other similar buildings owned by the same company. Indeed, the benchmarking process can be extended to individual technologies subcomponents.

Prudent users of benchmarking recognize that the choice of a benchmarking scope and metrics used for analysis can shape the conclusions drawn from the process [23]. By analogy, this can be readily seen in the case of vehicle fuel economy where during certain periods of time U.S. automobile efficiencies improved while vehicle miles traveled (also a metric that can be benchmarked) increased and per-vehicle fuel use thus remained constant and fleet-wide energy use rose. These four "competing" benchmarks are each accurate, but the choice of which of them to monitor can uniquely shape understanding and possible policy responses. As applied to peer-group definition, the performance of an automobile cannot be usefully compared to that of a mix of automobiles and large trucks. For buildings, filtering a heterogeneous data set even at a relatively high level of (e.g., schools versus hospitals) reveals large systematic differences in benchmarking domains.

An important caveat in benchmarking actual measured energy use (typically from utility bills or submeters) is that user "behavior" or other operational choices can influence the patterns observed as much as do physical attributes of the subject facility. Other confounding factors include year-to-year variations in weather and their influence on energy use. To continue with the vehicle analogy, official fuel-economy ratings are derived using a highly standardized test procedure, while the actual performance of a given vehicle often varies significantly from the standard value depending on loading, driving patterns, driving conditions, maintenance, etc. Such behavioral variations are one reason that energy intensity is not equivalent to intrinsic efficiency. An alternative to such "operational" ratings is to benchmark performance associated only with the buildings fixed elements (HVAC equipment, envelope, etc.). This is referred to as an "asset" rating, and must be performed using simulation. Asset-based methodologies are already deployed for homes (e.g., [16]) and are now under development for nonresidential buildings (e.g., [28]). Both approaches have value, and it must be kept in mind that, while eliminating various forms of noise from their assessments, asset-based techniques by definition do not capture the very real effects of building operations and management.

At one end of the buildings benchmarking spectrum lies whole-building benchmarking, in which all forms of energy and all end uses are aggregated into a single metric and compared against loosely similar types of buildings. The appeal of this approach is that it is conceptually simple and requires less time than other approaches. The limitation, however, is that less actionable information is yielded. While the relative performance of a given building may broadly suggest a potential to save energy, the specific pathways for doing so remain obscured.

At the other end of the spectrum, the most rigorous pathway to identifying applicable energy-efficiency measures is through in-depth energy audits and intensive simulation modeling. However, this is a costly proposition and requires considerable expertise. Midway between these extremes is an approach in which specific fuels and end uses are analyzed and logic applied in order to identify candidate energy-efficiency recommendations. This is known as "action-oriented" benchmarking [19] (Fig. 1). Action-oriented benchmarking improves on simplified benchmarking processes and helping lay the groundwork for investment-grade audits and professional engineering calculations.

In parallel with a benchmarking system's analytical underpinnings is the user interface through which users conduct the benchmarking process. More than a decade ago, Orlov et al. [20] reviewed the state of the art, including surveys of 22 early adopter companies that were using computer-based information dashboards (for a variety of purposes, outside the energy domain). They found that these systems were often "tentative and not linked to business processes" and contained "passive displays meant for executive eyes only." If dashboards are not connected to the people who "own" the processes they are evaluating, then the information does not become actionable. Similarly, metrics that do not fit the need are of little value, and can even be counterproductive. In sum, user-oriented design of benchmarking interfaces is essential if benchmarking methods are to be usefully adopted.
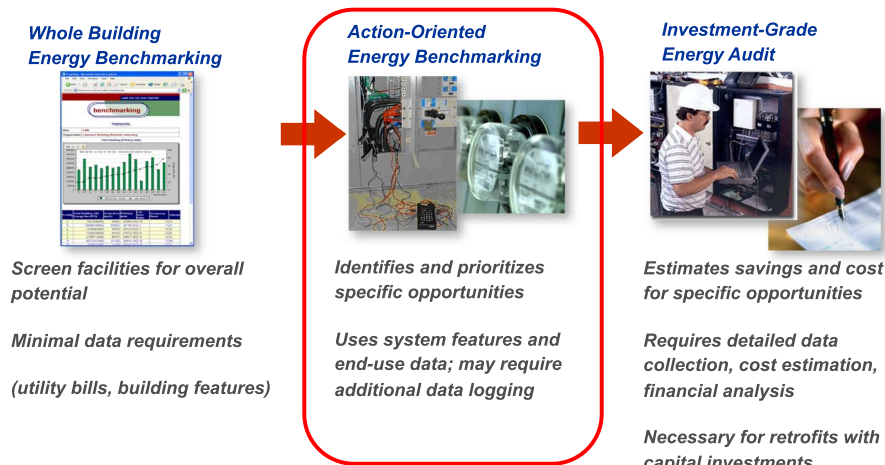
Fig. 1. *The spectrum of energy assessment approaches.*

II. BENCHMARKING PROGRAMS AND DRIVERS

There are several well-established benchmarking processes for nonresidential buildings. These include whole-building techniques focused on energy use such as the U.S. Environmental Protection Agency's ENERGY STAR Portfolio Manager for energy and water as well as point-based systems based on activities and attributes such as Leadership in Energy and Environmental Design (LEED), which covers a broad spectrum of green building attributes. As of 2015, over 25 000 buildings, representing more than 3.7 billion square feet of floor space of commercial building floor area had been benchmarked and certified using Portfolio Manager.[1] A subset of these have received an ENERGY STAR certification indicating that they perform in the top quartile of their peer group. As of mid-2013, approximately 44 000 U.S. buildings had been LEED rated, representing about 7 billion square feet of floor area.[2]

Federal and local directives mandating specific percentage reductions in overall energy use intensity (energy use per unit floor area) have also spurred the development of benchmarking processes. California Assembly Bill 1103[3] required electric and gas utilities to maintain energy consumption data for nonresidential buildings. Similarly, utility-bill disclosure requirements

have been established under California's AB531.[4] More recently, efforts have been increased in response to mandatory state and local energy disclosure laws that incorporate some form of benchmarking [9], [26]. These directives are often promulgated at the city level. Under the first year of Chicago's program, for example, energy use was publicly reported by 348 buildings representing 260 million square feet of space [5].

III. BENCHMARKING MECHANICS

Many building energy benchmarking procedures have been suggested, spanning a range of analytical techniques and data requirements [12]. One review identified 47 protocols for benchmarking nonresidential buildings and 31 that applied to residences [7].

At a high level, the methods fall into three broad categories, often used in combination:

- comparisons of actual measured data for a subject building to that of a cohort of buildings deemed "similar";
- use of characteristics and/or energy data to derive a unitless score or rating;
- specification of engineering-based simulation models or multivariate statistical models of a

---

[1]See http://www.energystar.gov/buildings/about-us/facts-and-stats.
[2]See http://www.usgbc.org/articles/infographic-leed-world.
[3]Government Code sections 11346.9(a). AB 1103 (Stats. 2007, ch. 533, §2), codified in pertinent part in Public Resources Code, section 25402.10, requires owners of nonresidential buildings to disclose to prospective buyers, lessees, and lenders the previous 12 months of the building's energy use in advance of the sale of the building, or the leasing or financing of the entire building, and to "benchmark" that data by providing a comparison of the building's energy use to that of other similar buildings.

[4]An act to amend Section 25402.10 of the Public Resources Code, relating to energy. Existing California law requires an owner or operator, on and after January 1, 2010, to disclose the U.S. Environmental Protection Agency's ENERGY STAR Portfolio Manager benchmarking data and rating to a prospective buyer, lessee of the entire building, or lender that would finance the entire building. The bill instead required the owner or operator to disclose the benchmarking data and rating to a prospective buyer, lessee of the entire building, or lender that would finance the entire building based on a schedule of compliance established by the State Energy Resources Conservation and Development Commission.

building to generate one or more performance estimates for comparison based on operational and/or climatic variables.

A key distinction here is that methods based on measured data intrinsically capture and incorporate impactful operational factors such as schedules and thermostat management, as well as the influences of fixed assets such as heating and cooling equipment efficiencies and the building envelope. They also incorporate what might be deemed "noise" such as fluctuations in occupancy, vacation periods, etc. Measures based on modeling standardize these operational factors or can otherwise separate or control for their influences. While standardizing or suppressing operational influences has a certain appeal, it also excludes factors that are intrinsic to a building's ultimate as-used performance and thus obscures opportunities to improve performance through better operations. Modeling complex nonresidential buildings requires substantial time and skill.

Of crosscutting importance in the benchmarking process is the assembly of a peer group of buildings against which to benchmark a subject building. Comparing the energy use of an office building against that of a diverse mix of buildings (offices, schools, hospitals, restaurants), for example, could easily lead to a distorted view of how the subject building is performing. More subtle considerations also come into play, for example the operating hours, geography, building size, or vintage. Comparing the performance of an office building in a hot-humid climate to an ostensible "peer group" in a cold-dry climate would not be meaningful. Peer-group definition is an ongoing area of research [6], as are methods for ensuring quality data, particularly when disparate sources are combined [2].

Fig. 2 illustrates the preceding points by presenting benchmark characteristics for the U.S. nonresidential building stock. The data derive from the U.S. Department of Energy's Commercial Buildings Energy Consumption Survey (CBECS) [27]. While the highest variation in energy intensities is seen when filtering by type of building, each differentiating factor correlates with statistically significant differences in EUI. Notably, the next greatest variation correlates with hours of occupancy, indicating that operational factors are just as important determinants of energy use as physical ones.

The California Commercial End Use Survey (CEUS) is perhaps the world's most thorough statistically representative localized repository of data for nonresidential buildings.[5] CEUS is a highly detailed survey of approximately 2800 nonresidential premises across California, based on a stratified random sampling across four utility regions, seven climate zones, and 62 building types. A standardized survey tool was used to document over 100 physical and operational characteristics of the building. In contrast to surveys relying on self-reporting, CEUS employed on-site surveys of building characteristics and monthly utility billing data. Short-term data logging and/or interval metering was performed at some sites and combined with calibrated simulation modeling to estimate end-use energy allocations and peak electrical demand. This high-quality data set enables a higher level of benchmarking granularity, ranging from campuses, to buildings, to systems, to components.

Fig. 3(a) and (b) illustrates heterogeneous peer groups that can be derived from the CEUS data set. Distinct peer group outcomes emerge not only for discrete building types but also when applying metrics (site/source energy,[6] cost, and emissions) to those subgroupings. This is a real-world reflection of fuel types as well as end-uses present (Fig. 4). It is readily apparent that median energy use intensities (EUIs) vary by building type. Hospitals and food service emerge as far more energy intensive given the greater abundance of special energy loads as well as more demanding HVAC requirements due to internal heat generation, airflow management requirements, etc. These distinctions, in turn, help analysts and building managers better target and prioritize areas to examine for energy-efficiency improvement. More energy-intensive buildings tend to offer a greater absolute savings potential, but targeting that potential requires benchmarking at the end-use level. It emerges clearly in Fig. 4, for example, that a much higher proportion of energy use in hospitals is attributable to space conditioning than in some other facility types, nearly 50% in this case, due primarily to high rates of once-through airflow.

The distinction between site and source energy is particularly important when exploring building performance at the end-use level, where carbon is a metric of interest, and where questions of which fuels to target are concerned. As suggested in Fig. 3(a) and (b), the differences in the absolute as well as relative energy intensities of varying building types can be more pronounced when considering site energy. This arises from varying levels of dependency on electricity versus other fuels.

An example of the influence these considerations can have on how a given subject building is "rated" against given peer groups is shown in Fig. 5, a case study of the California Energy Commission's headquarters. While the peer group sample size necessarily declines as stricter filters are applied, as the peer group becomes more aligned with the subject building the results become more meaningful. In this case, results for the most loosely defined peer group suggest that the subject building was not a particularly good performer.

---

[5] http://www.energy.ca.gov/ceus/ See also [10].

[6] "Site energy" counts the thermal value of energy at the point of consumption, whereas "Source energy" includes conversion losses at the power plant as well as through transmission and distribution. The differences for fuels (e.g., natural gas) are far less, reflecting only leakage in the distribution system.

Energy Use Intensity (kBtu/ft2-year, site energy): By type of building

By vintage

By geography

By heating equipment type
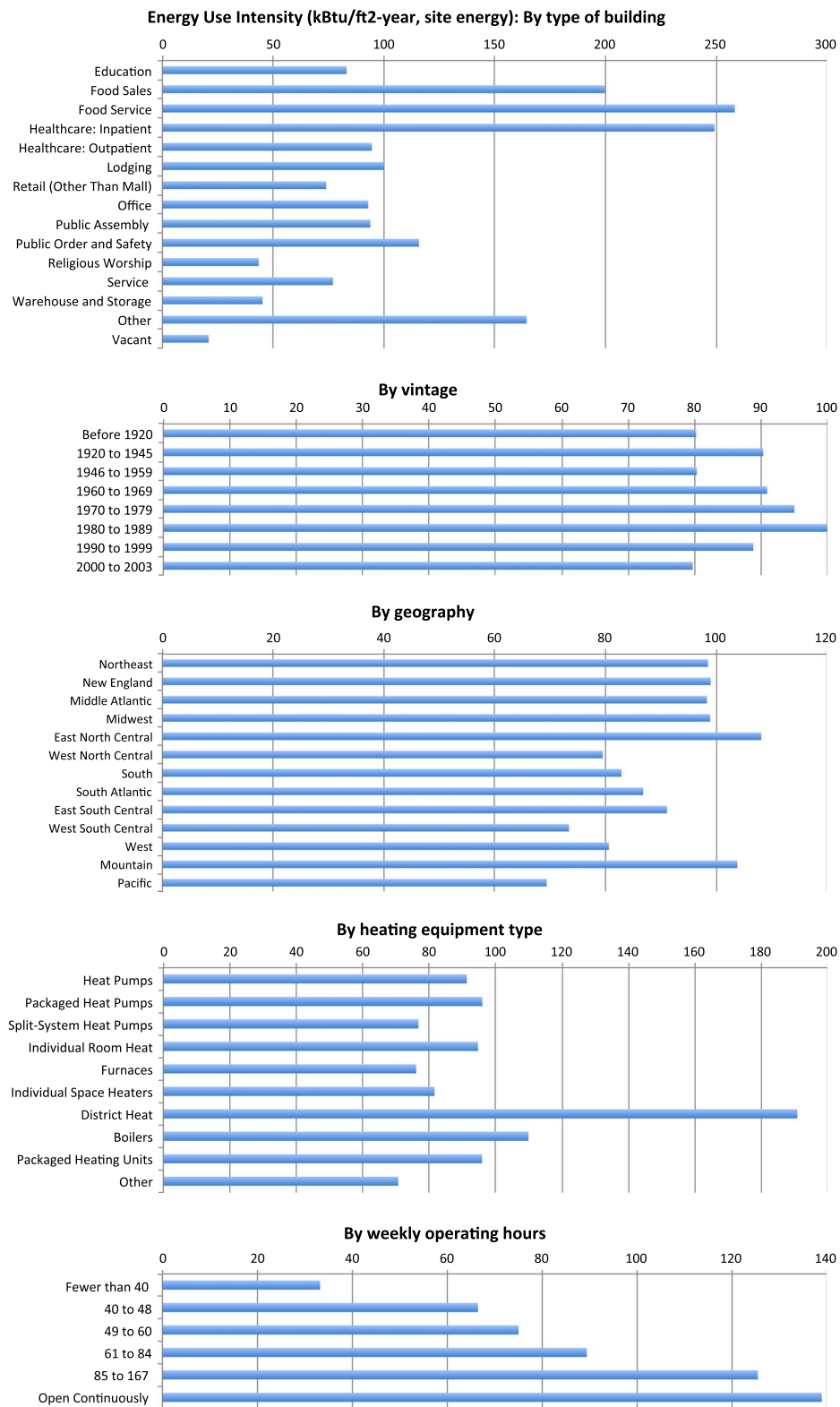
By weekly operating hours

Fig. 2. *Array of end use energy intensities for the U.S. nonresidential building stock (excluding mall buildings). Source: [27].*
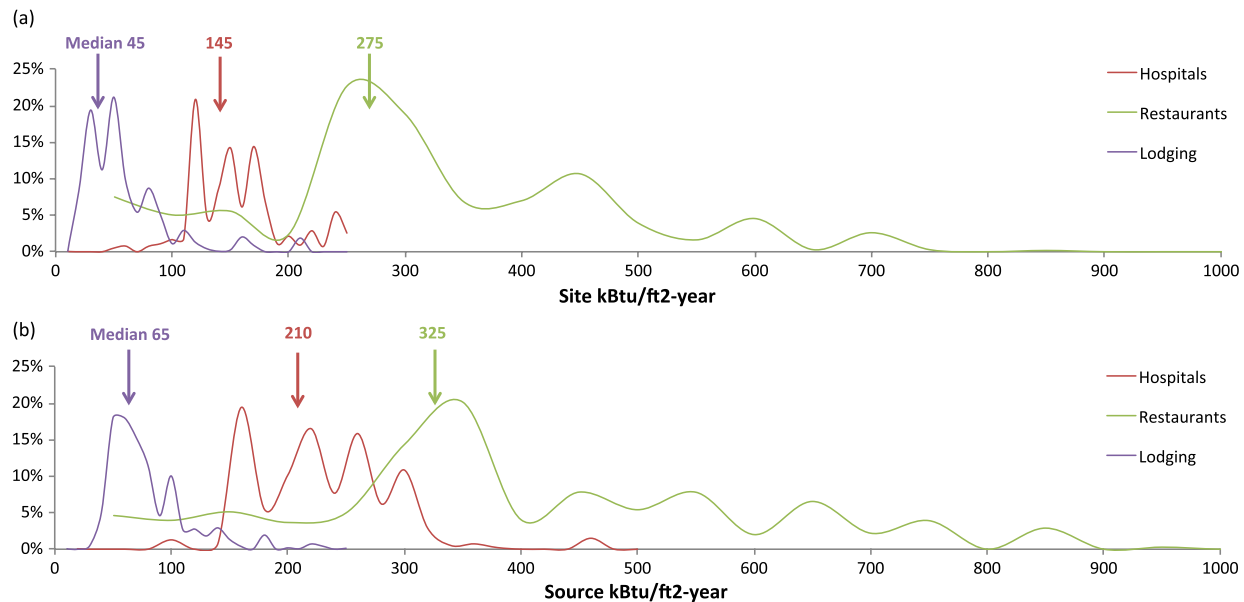
Fig. 3. *Frequency distribution of energy use intensity (EUI) by building type and site versus source energy. Filtering benchmark peer-group data sets by building type reveals differences in central tendencies and variance among cohorts. Source: [10].*

However, upon improving the filtering, relative performance improved considerably.

Peer groups can be derived from statistical surveys of a given building stock. Managers of real estate portfolios can also benchmark within their enterprise, rather than to a broader more heterogeneous population of buildings. Individual buildings can be "self-benchmarked" over time in order to track actual changes in performance. Definition of an appropriate peer group is context sensitive. For example, cohorts of buildings that are dominated by internal loads (e.g., grocery stores) need not be tightly limited to an individual climate or geography, whereas those that are more climate dominated (e.g., schools) should be organized accordingly.

Once a reasonable peer group is defined, and one or more filters applied to account for characteristics such as location, a benchmarking metric must be chosen and computed. The metric's numerator could be energy or some other factor of interest such as cost or greenhouse-gas emissions. The choice of denominator is important for normalization. While floor area is typically used, other factors may better characterize activities that occur in the building, such as meals served for a restaurant, bed-nights for a hotel, or students for a school.

The choice of benchmarking metrics is important. Even where floor area is used for normalization purposes, one can look at site or source energy or peak power, and at individual fuels versus whole-building energy use. As seen in Fig. 6, the choice of metrics can

yield qualitatively different conclusions in the case of restaurants. Here, for example, the high output of fast-food restaurants results in relatively high energy use per unit floor area, despite relatively low energy use per meal prepared.

Some facilities require specialized metrics. In the case of data centers, the ratio of total facility energy to the IT-related subset, known as the power usage effectiveness (PUE), is widely used. This metric provides important context on how efficiently the facility is cooled (the dominant use of energy besides the IT equipment itself). PUEs can range from 2 (one unit of cooling for each unit if of T "work") to just over 1 (negligible mechanical cooling required). In come cases, nonenergy benchmarks such as air-change rates (ACH) are highly meaningful. This is particularly true in the case of cleanrooms, where very substantial levels (up to 600 ACH per hour) are key drivers of overall facility energy use (Fig. 7). Even for subgroups with the same cleanliness rating, ACH varies widely. In turn, the required airflow can be provided with several different technologies. Fig. 8 relates the resulting airflow and air-movement choices to energy efficiency using a benchmark of cubic feet per minute of air movement per kilowatt of fan power input (CFM/kW). Looked at in this manner, this metric helps identify better practices for minimizing necessary airflow and the associated energy use.

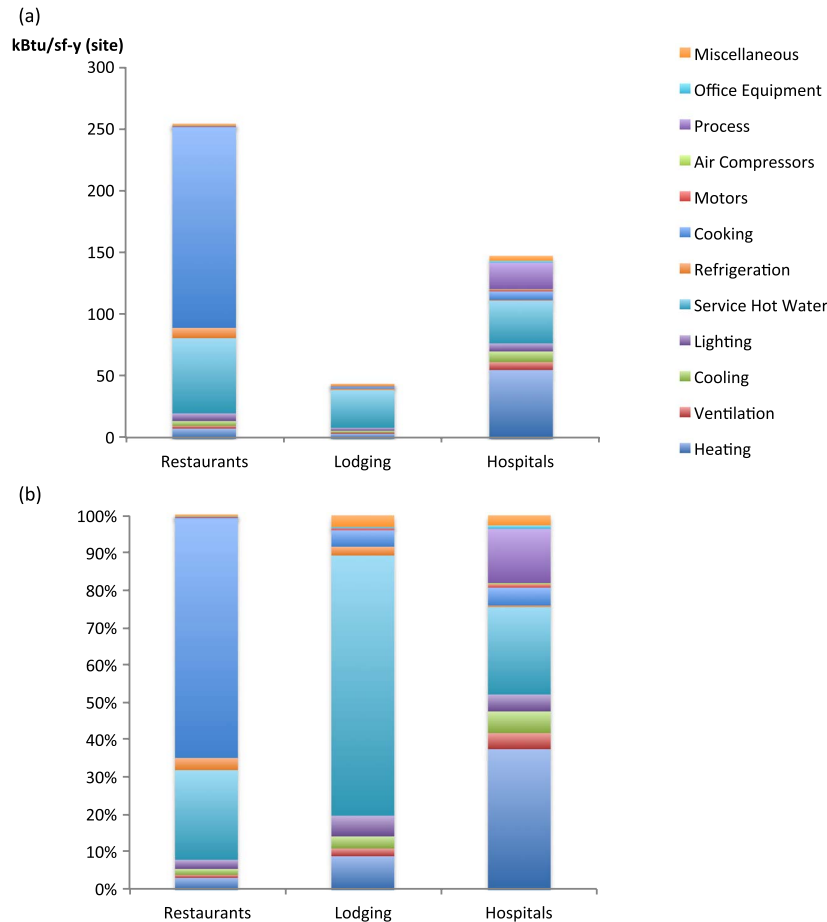Benchmarking can even be performed at the equipment level. For example, special-purpose benchmarks

Fig. 4. *In addition to variations in energy intensity (a), end-use shares (b) also differ widely. Benchmarking at the end-use level is thus an important part of the process of identifying efficiency opportunities. End uses in the diagram are in the same order as indicated in legend. Note that end uses are stacked in the order indicated by the legend. Source: [10].*

can be computed for high-performance computers such as those used for gaming and special effects. These can be extended even to subcomponents such as graphics-processor watts per unit of rendering performance, e.g., watts per frame-per-second. Wide variations in component efficiencies can be detected when benchmarking at this high level of specificity [15].

## IV. ACTION-ORIENTED BENCHMARKING

To achieve its full value, benchmarking should inform action. Whole-building benchmarks are highly constrained in this respect because they do not describe the determinants of particular energy outcomes. A layered approach, however, differentiating types of energy sources by end use, together with a profile of building characteristics and modes of operation begins to form the basis of analyses that can inform the identification of energy-efficiency opportunities.

In an illustration of the importance of granular information, Fig. 9 compares benchmark outcomes for elementary schools versus middle/secondary schools in California (using the CEUS data set). At the whole-building level, the ratio of site energy use per unit floor area is virtually the same, suggesting no difference between the two types of schools. When source energy is instead selected, elementary schools are found to be 6% less energy intensive. When individual fuels are isolated and benchmarked separately, it emerges that elementary schools are more fuel intensive but less electricity intensive. This explains the lower source-energy result and suggests that the secondary schools likely have more electric end-use devices. Moreover, elementary schools are found to have lower lighting energy intensities. When shifting to a per-student metric, the elementary schools are far less intensive than secondary schools, no doubt a reflection of more common areas per student (e.g., library, laboratory, gymnasium area) in secondary schools. Features benchmarks observed for these two
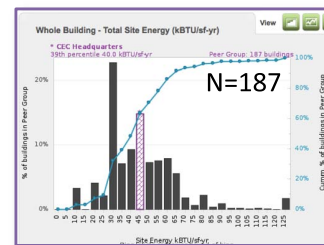
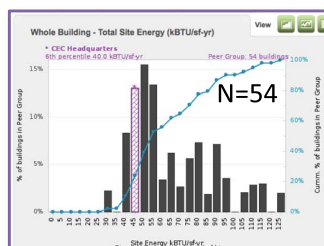Energy use of CEC Headquarters compared to: ...all California buildings
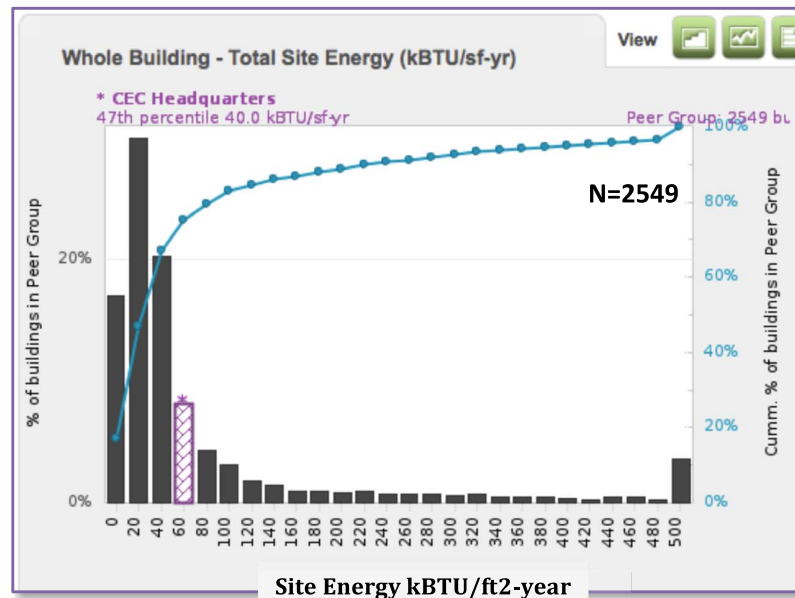
... only offices
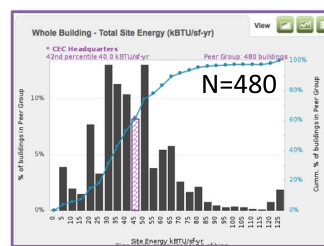
... + 1979-1990 vintage

... + 25-150k sf size range
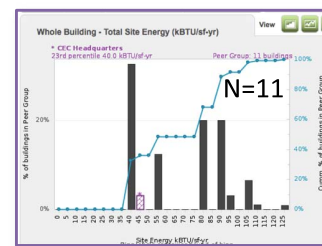

... + Central Valley

Fig. 5. *Analysis of California Energy Commission headquarters performed using EnergyIQ.lbl.gov. The distribution of peer-group energy intensities is represented by black bars, with the cumulative percentile indicated by the blue curve. The subject building's location in the peer-group spectrum for each case is indicated by a hatched purple bar. The five panels illustrate how relative benchmarking results for a given building can shift as the peer group is progressively defined. In this case, the peer group sample size declines, by definition, but its relevance increases. The subject building is found to be more energy intensive than average when compared to all building types, vintages, and sizes in all California locations (first panel), but improves to among the more efficient buildings within a progressively homogenous peer group.*

types of schools also reveal that cooling systems are on average five years older in secondary schools, an important difference given the continued tightening of efficiency standards for such equipment, together with performance deterioration with age.

The ability to benchmark at the level of individual end uses, together with the application of a variety of metrics, can help identify specific energy-efficiency improvement opportunities, as suggested in Fig. 10.
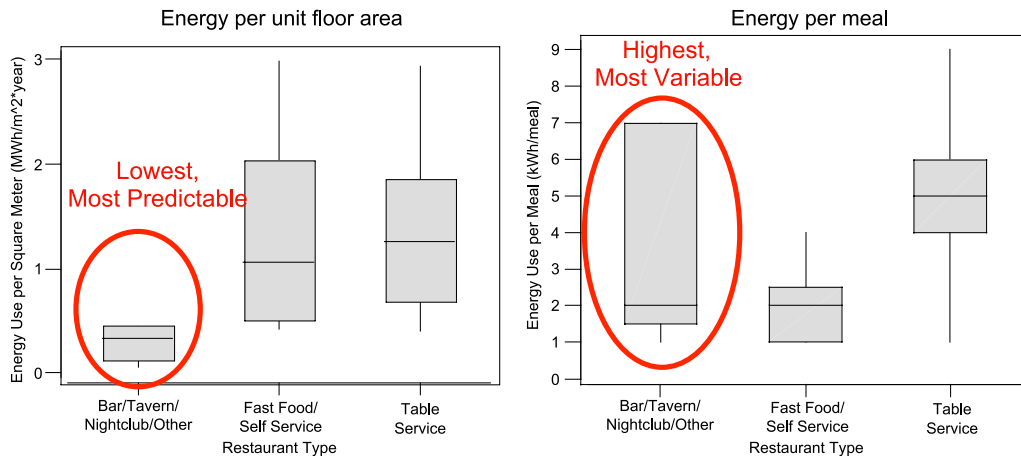
Fig. 6. *Benchmarks of several types of food service facilities, based on different metrics. Shown are statistically representative peer groups within the state of California. Source: 1999 California Commercial End Use Survey.*

V. ASSESSING USER NEEDS

To gain insight into the functionality desired by potential users of action-oriented benchmarking tools, we distributed a survey to 500 stakeholders around the United States [13]. The 95 respondents collectively influence 554 million square feet of building floor area (either as owners, occupants or service providers). The key results were as follows.

- Almost three-quarters of the respondents already utilize some sort of energy benchmarking process, applying it to a very wide range of building types.

- One in five respondents conduct some form of nonenergy benchmarking (e.g., sales per employee), which suggests an opportunity to add value by enabling a benchmarking tool to utilize the same normalization factors.

- The three main reasons given for buildings energy benchmarking were identifying energy-efficiency opportunities, prioritizing investments, and making comparisons to other facilities. A quarter of respondents provided additional reasons, such as verifying energy savings, tracking persistence of savings, and making the business case to management for efficiency investments.
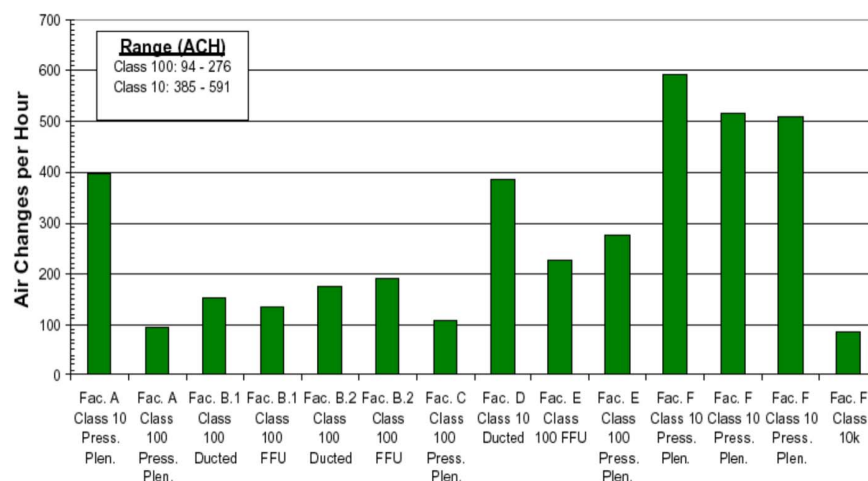
Fig. 7. *Air circulation is one of the prime drivers of energy use in cleanrooms. A key benchmark metric in this regard is air changes per hour. Cleanroom particle count ("Class") is indicated in bar labels, with 10 being the lowest count and 10 000 being the highest. In the figure it can be seen that some cleanrooms achieve target particle removal levels with lower air change rates (a strong correlate of energy intensity). Source: [24].*
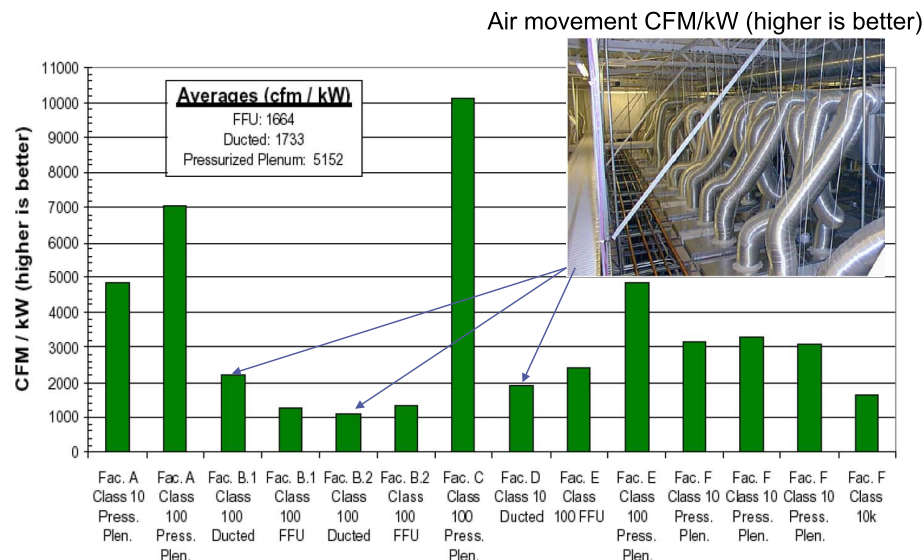
Air movement CFM/kW (higher is better)

Fig. 8. *Within cleanroom air circulation systems, the rate of air movement per unit of fan power is an important efficiency metric. In the figure it can be seen that among pressurized plenum designs energy efficiency can be several times that of other strategies (up to 10 000 cubic feet per minute per kW of fan energy input). Among the least efficient are ducted systems (photo insert), which experience significant losses due to pressure drop. Source: [24].*

- Users assigned particularly high importance to six types of metrics: whole building, end use, peak power, energy cost, emissions, and productivity (e.g., energy cost per customer). Equal importance was placed on applicability to benchmark new versus existing buildings and to perform cross-sectional versus longitudinal benchmarking. The ability to exchange data among benchmarking tools was also assigned a high importance.

Fig. 9. *Relative performance of elementary and secondary schools in California as a function of benchmark metric. See footnote 6 for definitions of site and source energy. Source: CEUS database.*

- Users desired to be able to include other users' benchmarking results in the peer groups to which they compare themselves.
- Respondents fell into two cohorts with respect to the time they are willing to spend using a benchmarking tool (Fig. 11). One group centered on the 0–60-min range while the other in the vicinity of 120 min or more. This bimodal pattern held across all user types (e.g., owners, tenants, service providers).
- Virtually all respondents desire both graphical and tabular outputs. Only seven percent wanted graphics only and only one percent wanted tables only.
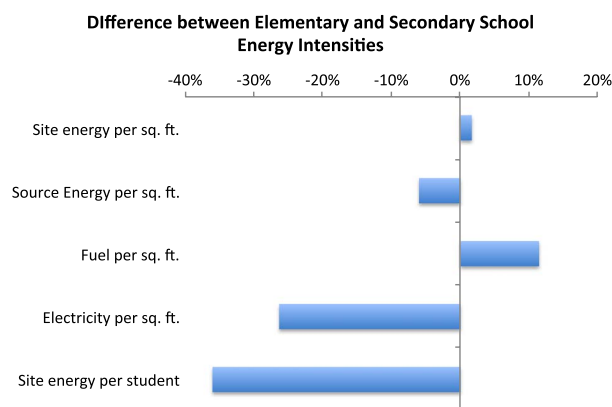
The work of ASHRAE Technical Research Project 1286 offered another assessment of best practices for energy benchmarking tool design [7]. Their findings indicated significant room for improvement in terms of analytical capabilities of benchmarking tools, as well as usability of the interfaces. A particular need identified was for tools that could be used to assess efficiency opportunities and recommend specific energy-efficiency "actions."

VI. ENERGYIQ—A WEB-BASED TOOL FOR ACTION-ORIENTED BENCHMARKING

In isolation, benchmarking can inspire action but provides no practical guidance. With sponsorship from the California Energy Commission's Public Interest Energy Research (PIER) program and the California Air Resources
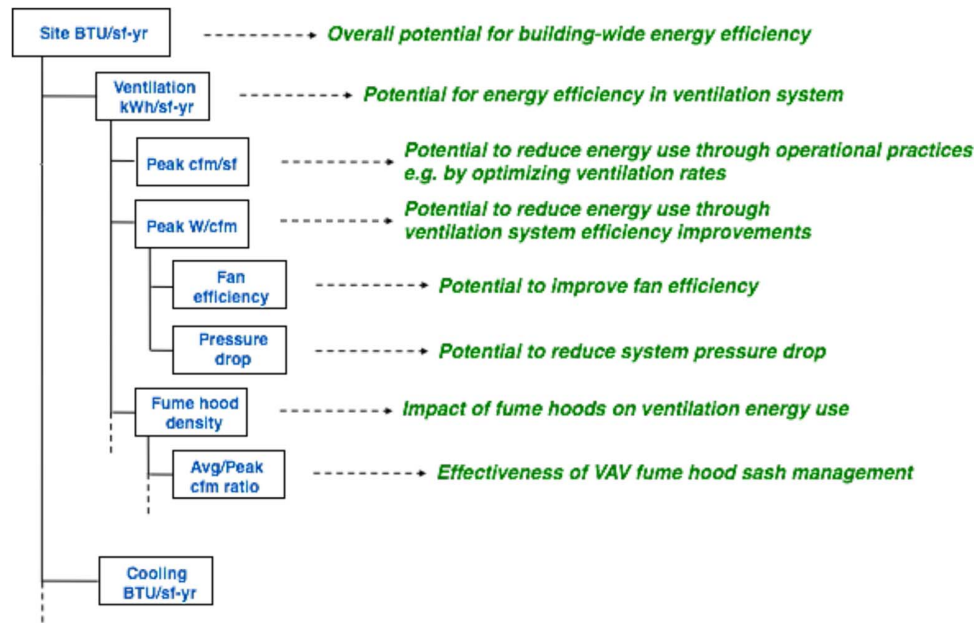
Fig. 10. *Conceptual illustration of how conventional "whole building" benchmarks such as site BTUs/ft2-year can become progressively informative when they are focused at the end-use level (in this case ventilation) together with specialized metrics that pertain to specific technology opportunities.*

Board, the U.S. Department of Energy's Lawrence Berkeley National Laboratory constructed the next generation of energy benchmarking methods to address this problem. EnergyIQ[7]—the first "action-oriented" benchmarking tool for nonresidential buildings—provides a standardized opportunity assessments based on benchmarking results, along with decision-support information to help refine action plans [13] (Fig. 12).

EnergyIQ is cloud based to facilitate access, transparently documented, and available at no cost to users. The tool benchmarks energy use, costs, and features for all major building types and provides a carbon-emissions calculation for the energy consumed in the building, an important part of any businesses' overall carbon footprint. Application programming interfaces (APIs) allow the underlying data and benchmarking engine to be used in any website [18]. The system interoperates with the ENERGY STAR Portfolio Manager, and allowing those users to extend their whole-building assessment without the need for dual data entry. Actual as well as modeled energy data for an individual building can be input into the tool.

The design of EnergyIQ was informed by the aforementioned target-audience research, as well as the findings and recommendations of ASHRAE Research Project 1286. The tool does not underpin any particular program or campaign, and does not provide a pass/fail score.

[7]http://energyiq.lbl.gov

The aforementioned CEUS database provided the initial California-centered peer-group data that underlay the benchmarking process. A national data set, based on the 2003 Commercial Buildings Energy Consumption Survey (CBECS) provides an alternative reference point for benchmarking in any U.S. location outside California. CBECS is a statistically representative sample of 5215 buildings across the country. Simulation models of each CEUS building are calibrated to actual utility bills [10] and then used to estimate end-use energy splits and to evaluate energy savings opportunities. EnergyIQ users can also benchmark exclusively against their own building portfolio, or against buildings entered by other users. Peer-group data are provided as distributions (rather than point estimates). All data (peer group as well as a user's own building) are anonymized.

EnergyIQ also uniquely employs what we refer to as "features benchmarking." The premise is that there is value in benchmarking the presence or absence of certain features in a quantitative or qualitative fashion. Employing the features-benchmarking module enables users to see distributions of energy-related characteristics among the peer-group buildings. Approximate 85 energy-related characteristics can be analyzed (a subset of those collected in the original CEUS surveys), a diversity of which can be viewed, such as lighting type, HVAC equipment, and plug loads. Features benchmarking includes metrics such as equipment efficiencies (e.g., kW/ton)
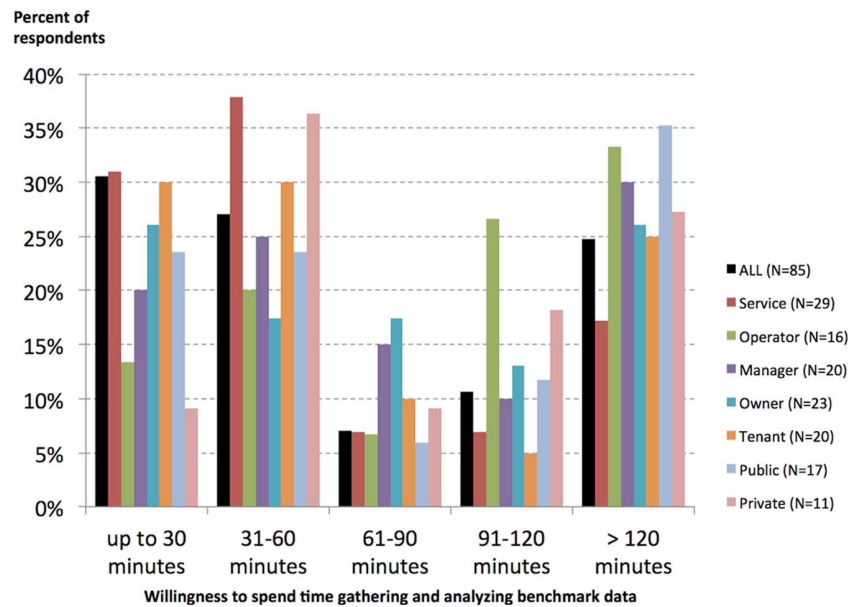
Fig. 11. *Bimodal distribution of amounts of time prospective users of benchmarking tools are willing to spend gathering and entering data for a single benchmarking session. Some expected variations are visible, e.g., that tenants seek to spend less time than operators. Number of respondents indicated in legend.*

and product categories (e.g., chiller types). Features benchmarking also distinguishes among important operational modes such as types of space-conditioning controls, temperature settings, and hours of occupancy.

The tool speeds the path to usable results by allowing the user to browse a wide variety of metrics and visualizations [e.g., as in Fig. 12(b)] generated dynamically based on peer-group data via the web interface. The user can enter the data necessary to map their own building onto a given visualization. This contributes to the "action-oriented" philosophy of the tool requiring the user to enter only the data necessary to obtain the analysis they seek and metrics that have meaning for their particular situation. Users can stipulate performance targets and track progress toward them [Fig. 12(c)].

The benchmarking process itself is highly customizable, offering extensive filters. Users specify metrics of their choice, in terms of energy quantities, costs, or greenhouse-gas emissions (SI and British units). Energy-related data visualizations and metrics include total energy use, electricity, or fuel, each at the whole building and end-use level. Peak demand is also an optional metric, and one not typically included in benchmarking tools. In addition to floor-area-based energy intensities, EnergyIQ metrics normalize consumption by factors such as per employee for any building type, per seat for food service, per student for schools, per patient beds for hospitals, and per guest room for lodging building types.

Four general categories of graphical presentation are used: simple summaries such as tables, frequency distributions (quartiles, ranked, histogram, or scatter diagrams), and conventional bar charts visualizing indicators such as equipment efficiency. Longitudinal (multiyear) visualizations are also available, enabling a user to track evolving benchmarks over time. Ultimate metrics can thus take many forms, ranging from traditional values such as total energy use per square foot per year to alternatives such as lighting-related pounds of CO_2 per restaurant seat per year.

For California buildings, EnergyIQ offers the capability to perform an "upgrade analysis" to examine the impact of implementing a select group of 35 distinct energy-efficiency measures across ten end-use categories, with 132 efficiency-level variations [Table 1 and Fig. 12(d)].[8] The outputs are organized into three categories: the likely relevance or "fit" of a particular energy saving measure, significance of potential savings, and cost effectiveness. The methodology employs the subset of CEUS buildings that match the user's peer group, and then presents the range of savings for those peer buildings based on the implementation of those measures within hourly eQUEST buildings energy simulation models for each building in the underlying peer group [8]. The underlying parametric simulation database includes over 65 000 measure-building combinations. The measures are, by definition, only applied to those buildings for which they are applicable (e.g., compact-fluorescent lighting only to those buildings

---

[8]See EnergyIQ technical documentation: https://sites.google.com/a/lbl.gov/energyiq/methodology/upgrade-analysis.
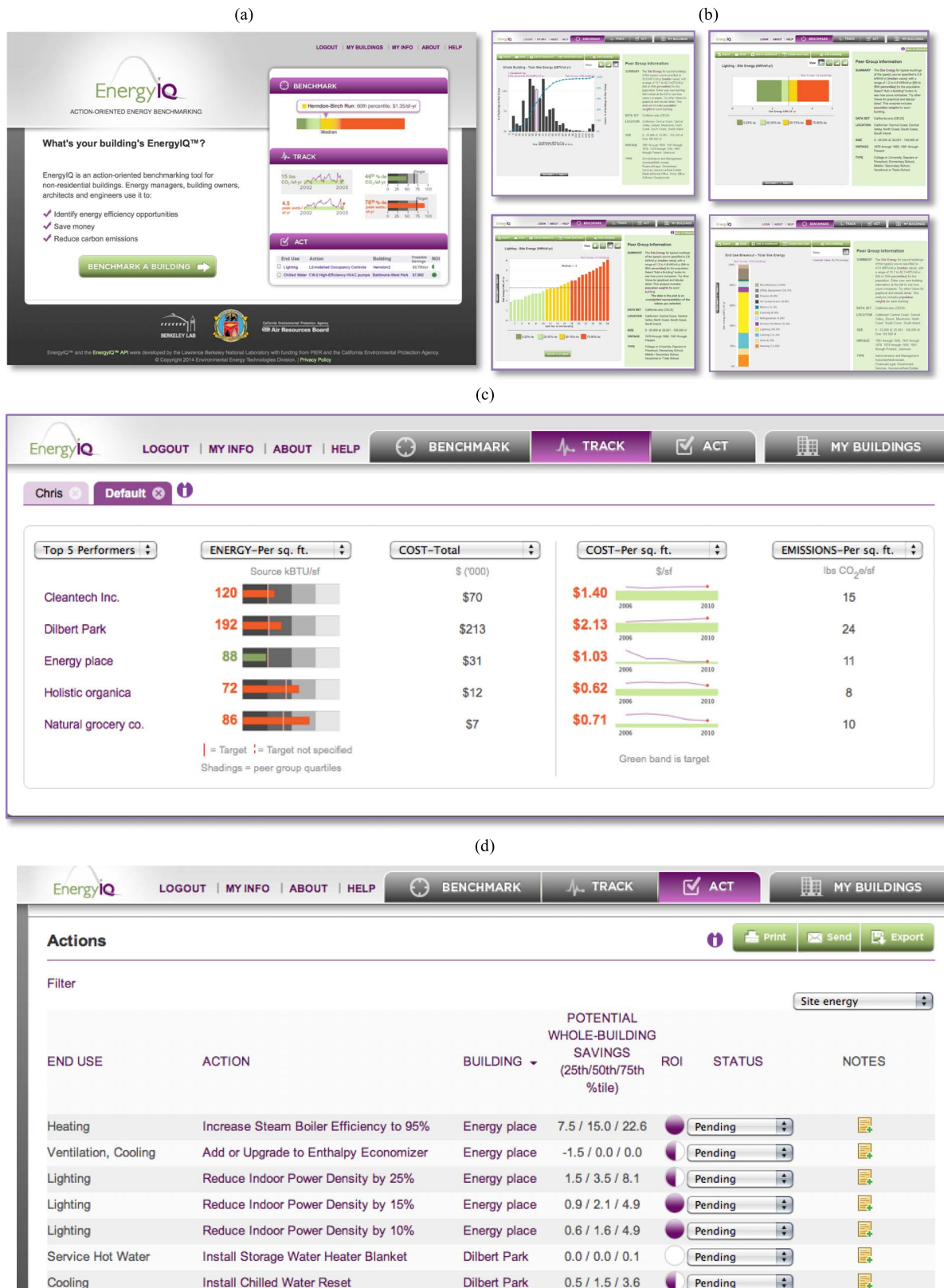
(a)

(b)

(c)

(d)

Fig. 12. *Primary EnergyIQ user data visualizations. (a) EnergyIQ homepage. (b) Data visualizations. (c) Dashboard: portfolio-level assessment by metric. (d) Actions: savings opportunities by building, end-use, and measure.*

Table 1 Types of Measures Evaluated by EnergyIQ to Determine Potential Efficiency Opportunities

End use	Types of measures	Number of measures	End use	Types of measures	Number of measures
Envelope	Roof insulation	1	Refrigeration	Compressor: efficiency	2
	Radiant barriers	1		Subcooling: ambient	1
	High-albedo roof	1		Subcooling: mechanical	1
	Wall insulation	5		Condensor type	1
	Install more efficient windows and frames	4		Condensor sizing	1
Outdoor lighting	Photocell control	1	Packaged HVAC	Controls	1
	Compact fluorescent lamps	5		High efficiency DX units: efficiency and type	42
	Linear fluorescent lamps	2		Upgrade to enthalpy type economizer	1
	HID lamps	1	Built-up HVAC	Reduce VAV air flow	4
Indoor lighting	Reduce lighting power density	3		Constant- to variable-volume system	1
	Compact fluorescent lamps	5		Convert electric reheat to gas boiler	1
	Linear fluorescent lamps	2		Upgrade to enthalpy type economizer	1
	HID lamps	1	HVAC chillers	Improved efficiency	24
Service hot water	Storage tank insulation	1		Chilled water reset	1
	High efficiency water heaters	3	HVAC boilers	Improved efficiency	6
	High efficiency boilers	3	HVAC motors	Improve efficiency pumps	1
	Pipe insulation	1		Variable speed pumps	1
				Two-speed pumps	1
				Efficient fan motors	1

using incandescent lighting). Note that when computing whole-building metrics, HVAC interactions are included. Waste heat emanating from loads such as lighting or plug loads will be delivered to the conditioned space, reducing heating requirements (and thus potential savings) in winter and increasing cooling requirements in summer. Outside California, lists of applicable efficiency improvements are generated but without the quantitative assessment made possible by the California-specific CEUS data.

As of 2015, there were approximately 1400 users of the tool who had collectively evaluated 140 million square feet of building floor area (over 1000 buildings). Users range from individual building owners, to intermediaries such as property management companies, to third-party energy management stakeholders such as utilities and energy service companies.

VII. IMPACTS OF BENCHMARKING

Attributing marketplace decisions to implement energy-efficiency measures to the use of benchmarking is an elusive goal at best. However, while improved information does not in and of itself achieve energy-efficiency improvements, it is critically enabling. A more nuanced view is that benchmarking enables the identification and ranking of opportunities, creates awareness and attention, and provides intelligence that enables building operators to remain vigilant and ensure that intended performance targets are met and persist over time. Ongoing benchmarking also builds confidence about performance levels and savings claims, hence managing investment risks. Recent efforts have helped to formalize the process of evaluating the impact of benchmarking and disclosing benchmark data in the marketplace [26].

The developers of Portfolio Manager have observed that the 35 000 buildings receiving ENERGY STAR scores over the period 2008–2011 reduced their aggregate energy use at an average rate of 2.4% per year [29]. For the three-year period in question, aggregate savings ranged from 2% (hospitals) to 11% (retail stores).

There has been considerable controversy around efforts to assess the energy savings resulting from the LEED benchmarking and rating process. An early

publication asserting savings of 25%–30% [25] failed to apply rigorous peer-group standardization, yielding incomparable data sets of LEED and non-LEED buildings. The study also received criticism for not weighting the multibuilding results by floor area, conflating median and mean results, and failing to conduct statistical significance tests on the outcomes. Reanalysis suggested site energy savings of 10% across various commercial building types (17% for offices), but no source energy savings [21], [22]. Buildings with the most stringent LEED ratings (Gold and Platinum) had the lowest energy use. Following these criticisms, the LEED methodology required reporting of measured energy use and incorporated the Portfolio Manager method of energy analysis. It should be kept in mind that energy is just one of many "green" attributes for which points are earned toward a LEED rating.

In contrast to the aforementioned highly aggregated whole-building benchmarks, action-oriented benchmarking procedures help pinpoint specific sources of energy inefficiency in a highly verifiable fashion. In one case study, an extensive system installed at the University of California Merced campus monitors 3000 data points per building, polled every 15 min [14]. An overarching benchmarking protocol along with focused data visualization enabled specific issues to be identified. In one example, wide variations in benchmarked fan efficiencies revealed miscalibrated sensors, which, upon correction, yielded 30% improvement in fan performance. In another example, benchmarked zonal differences between setpoint and actual temperatures identified previously and enabled the correction of previously undetected deficiencies in one of nine variable air volume boxes. These examples illustrate the importance of end-use benchmarking and features-based benchmarking methods for identifying efficiency opportunities.

VIII. CONCLUSION AND BENCHMARKING FRONTIERS

Buildings energy benchmarking has evolved from simplified whole-building methods to highly disaggregated approaches that capture important nuances and provide the ability to assess energy saving opportunities at relatively low cost. Extensive peer-group data coupled with relevant and compelling metrics are essential components of a usable benchmarking protocol. Operational and asset assessment strategies all have strengths and weaknesses.

The use of benchmarking will continue to be stimulated by public policies such as utility disclosure ordinances. For example, Chicago recently compelled all large buildings to perform and disclose Portfolio Manager ratings, revealing that 46% of that building stock fell short of the ENERGY STAR performance target [5].

Many frontiers remain to be explored. More comprehensive and regularly updated building energy surveys are needed in order to produce the statistically representative peer group data needed for benchmarking. The CEUS and CBECS data are already more than a decade old. Measured end-use data are particularly rare. The advent of public-domain "big data" and its application to the buildings energy arena through projects such as the Buildings Performance Database [2][9] is yielding new sources of peer-group data and larger data sets that promise to enable more fine-grain filtering than is currently possible. In some areas (where disclosure is mandatory) 100% samples of buildings are becoming available.

The integration of energy data acquisition, visualization, and building management systems has been pursued for some time. Much more can be done to integrate benchmarking into the process of operating buildings and diagnosing deficiencies that lead to energy waste. For example, the value of building automation and diagnostic systems would be enhanced by embedding benchmarking functionality. However, to be more broadly adopted, benchmarking user interfaces must be designed with target users in mind, with an emphasis on usability.

Benchmarking techniques must keep up with continually emerging technologies for energy management. Demand-response technologies, vehicle-to-building systems, and the trend toward zero net energy buildings all call for new types of metrics and benchmarking protocols. The now wide diffusion of interval metering has created the possibility of benchmarking building performance at much finer time steps, and the use of these data for diagnosing new sorts of efficiency and load-management opportunities.

While tools such as EnergyIQ have begun to identify energy-efficiency opportunities based on benchmark results, they do not include cost-benefit assessments. This is challenging given the highly variable economic environment in which energy upgrades are implemented. While upgrade costs are often stipulated for residential buildings, the scope for meaningfully doing so for nonresidential buildings is far more limited.

Benchmarking exists within a broader process of improving energy use in buildings, including the conduct of investment-grade energy audits and the development of work specifications for actually performing energy saving improvements. More research is needed to determine the impact and market value of benchmarking in this broader context.

Benchmarking is today most commonly performed by energy auditors and commissioning agents, but much scope remains to further integrate benchmarking results into the broader energy information environment. Examples include enhancing utility bills to provide building occupants with a recurring contextual evaluation of how their performance compares to others. Other market transactions that could be enhanced by energy

---

[9]http://energy.gov/eere/buildings/building-performance-database

benchmarking include real estate acquisition, finance, and appraisal [1]. Insurance companies are also increasingly engaging in the green buildings movement [17] and require metrics in order to determine whether a subject building may be eligible for special products and services.

One of the greatest challenges facing the practice of energy benchmarking is that it does not fit neatly into the set of energy-efficiency strategies often recognized and rewarded through voluntary energy-efficiency programs and policies such as utility rebate programs and

building codes. Traditionally, financial incentives for energy efficiency are awarded largely to discrete technologies ("wigits") rather than to enterprise-level practices that improve performance in a crosscutting manner. This is a known challenge that merits consideration by energy regulators and policymakers. Incentivizing building energy benchmarking is a key strategy for facilitating the identification of applicable technologies, and ensuring that their intended performance is achieved in practice. ∎

REFERENCES

[1] M. Black et al., "Valuation of green and high performance properties: Background and core competency," The Appraisal Foundation, Washington, DC, USA, APB Valuation Advisory #6, 2015, p. 46

[2] R. E. Brown, T. Walter, L. N. Dunn, C. Y. Custodio, and P. A. Mathew, "Getting real with energy data: Using the buildings performance database to support data-driven analyses and decision-making," in Proc. ACEEE Summer Study Energy Efficiency in Buildings, 2014, pp. 11-49–11-60.

[3] M. V. Chester, A. Horvath, and S. Madant, "Comparison of life-cycle energy and emissions footprints of a passenger transportation in metropolitan regions," Atmospheric Environ., vol. 44, pp. 1071–1079, 2010.

[4] W. Chung, "Review of building energy-use performance benchmarking methodologies," Appl. Energy, vol. 88, pp. 1470–1479, 2011.

[5] City of Chicago, "2014 city of Chicago building energy benchmarking report," 2014, p 29. [Online]. Available: http://www.cityofchicago.org/energybenchmarking

[6] X. Gao and A. Malkawi, "A new methodology for building energy performance benchmarking: AN approach based on intelligent clustering algorithm," Energy Buildings, vol. 84, pp. 607–616, 2014.

[7] J. Glazer, "Evaluation of building energy performance rating protocols," Prepared for ASHRAE Technical Committee 7.6—Systems Energy Utilization, Dec. 6, 2006, p. 255.

[8] J. Hirsh, "eQUEST: The quick energy simulation tool," p. 8, 2009. [Online]. Available: http://www.doe2.com/equest/

[9] Institute for Market Transformation (IMT)/ CBRE, "Guide to state and local energy performance ratings: Version 3," 2013. [Online]. Available: http://www.imt.org/uploads/resources/files/GuidetoStateandLocalEnergyPerformanceRegulations3.pdf

[10] Itron, "California end use survey," Prepared for the California Energy Commission, Rep. CEC-400-2006-005, 2006. [Online]. Available: http://www.energy.ca.gov/2006publications/CEC-400-2006-005/CEC-400-2006-005.PDF

[11] J. Ke, L. Price, M. McNeil, N. Z. Khanna, and N. Zhou, "Analysis and practices of energy benchmarking for industry from the perspective of systems engineering," Energy, vol. 54, pp. 32–44, 2013.

[12] Z. Li, Y. Han, and P. Xu, "Methods for benchmarking building energy consumption against its past or intended performance: An overview," Appl. Energy, vol. 124, pp. 325–334, 2014.

[13] P. Mathew, E. Mills, N. Bourassa, and M. Brook, "Action-oriented benchmarking: Using the CEUS database to benchmark commercial buildings in california," Energy Eng., vol. 105, no. 5, pp. 6–18, 2008.

[14] A. Mercado and J. Elliott, "Energy performance platform: Revealing and maintaining energy efficiency with a customized energy information system," in Proc. ACEEE Summer Study Energy Efficiency Buildings, 2012, pp. 12–166, 12–178.

[15] N. Mills and E. Mills, "Taming the energy use of gaming computers," Energy Efficiency, 2015, pp. 1–18.

[16] E. Mills et al., "Asset rating with the home energy scoring tool," Energy Buildings, vol. 80, p. 441–450, 2014.

[17] E. Mills, "The greening of insurance," Science, vol. 338, p. 1424, 2012.

[18] E. Mills and P. Mathew, "Web services that foster innovation in buildings energy analysis tools," in Proc. ACEEE Summer Study Energy Efficiency Buildings, 2012, pp. 12-179–12-191.

[19] E. Mills, P. Mathew, N. Bourassa, M. Brook, and M. A. Piette, "Action-oriented benchmarking: Concepts and tools," Energy Eng., vol. 105, no. 4, pp. 21–40, 2008.

[20] L. Orlov, C. Mines, L. Herbert, and C. Teubner, "Making dashboards actionable," Forrester Res., pp. 1–18, 2003.

[21] J. H. Scofield, "No evidence LEED building certification is saving primary energy," APS News, vol. 22, no. 7, 2013. [Online]. Available: http://www.aps.org/publications/apsnews/201307/backpage.cfm

[22] J. H. Scofield, "A re-examination of the NBI LEED building energy consumption study," in Proc. Energy Program Eval. Conf., Portland, OR, USA, 2009, pp. 764–777.

[23] M. Sisk, "Are the wrong metrics driving your strategy?" Harvard Manage. Update, 2003, Art. No. U0311A.

[24] W. Tschudi and T. Xu, "Cleanroom energy benchmarking results," ASHRAE Trans., no. Winter, pp. 733–742, 2003.

[25] C. Turner and M. Frankel, "Energy performance of LEED for new construction buildings," New Buildings Institute, White Salmon, WA, USA, Final Rep., 2008.

[26] U.S. Dept. Energy, "Benchmarking & transparency policy and program impact evaluation handbook," Prepared by Navigant Consulting, Inc. and Steven Winter Associates, Inc., Rep. DOE/EE-1220, 2015, p. 86.

[27] U.S. Dept. Energy, "Commercial buildings energy consumption survey (CBECS) 2003," 2015. [Online]. Available: http://www.eia.gov/consumption/commercial/data/2012/

[28] U.S. Dept. Energy, "Building energy asset score," 2014. [Online]. Available: http://energy.gov/eere/buildings/building-energy-asset-score

[29] U.S. Environmental Protection Agency, "Benchmarking and energy savings," Energy Star Program Rep., 2012. [Online]. Available: www.energystar.gov/ia/business/downloads/datatrends/DataTrends_Savings_20121002.pdf?8d81-8322

ABOUT THE AUTHOR

Evan Mills received the Ph.D. degree in environmental and energy systems studies from the Faculty of Physics, Institute of Technology, Lund University, Lund, Sweden, in 1991.

He is a Senior Scientist at the Lawrence Berkeley National Laboratory, U.S. Department of Energy, Berkeley, CA, USA. His research spans the domains of energy management, risk management, and climate change impact assessment. He has published nearly 300 technical articles and reports and has contributed to 11 books. He is a participant in the Intergovernmental Panel on Climate Change (IPCC), which shared the 2007 Nobel Peace Prize.